



Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis

Jensen, Dan Børge; Hogeveen, Henk; De Vries, Albert

Published in:
Journal of Dairy Science

DOI:
[10.3168/jds.2015-10060](https://doi.org/10.3168/jds.2015-10060)

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY-NC-ND](#)

Citation for published version (APA):
Jensen, D. B., Hogeveen, H., & De Vries, A. (2016). Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. *Journal of Dairy Science*, 99(9), 7344-7361. <https://doi.org/10.3168/jds.2015-10060>



Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis

Dan B. Jensen,^{*1} Henk Hogeveen,[†] and Albert De Vries[‡]

^{*}Department of Large Animal Sciences, University of Copenhagen, DK-1870 Frederiksberg C, Denmark

[†]Chair Group Business Economics, Wageningen University, 6706 KN Wageningen, the Netherlands

[‡]Department of Animal Sciences, University of Florida, Gainesville 32611

ABSTRACT

Rapid detection of dairy cow mastitis is important so corrective action can be taken as soon as possible. Automatically collected sensor data used to monitor the performance and the health state of the cow could be useful for rapid detection of mastitis while reducing the labor needs for monitoring. The state of the art in combining sensor data to predict clinical mastitis still does not perform well enough to be applied in practice. Our objective was to combine a multivariate dynamic linear model (DLM) with a naïve Bayesian classifier (NBC) in a novel method using sensor and nonsensor data to detect clinical cases of mastitis. We also evaluated reductions in the number of sensors for detecting mastitis. With the DLM, we co-modeled 7 sources of sensor data (milk yield, fat, protein, lactose, conductivity, blood, body weight) collected at each milking for individual cows to produce one-step-ahead forecasts for each sensor. The observations were subsequently categorized according to the errors of the forecasted values and the estimated forecast variance. The categorized sensor data were combined with other data pertaining to the cow (week in milk, parity, mastitis history, somatic cell count category, and season) using Bayes' theorem, which produced a combined probability of the cow having clinical mastitis. If this probability was above a set threshold, the cow was classified as mastitis positive. To illustrate the performance of our method, we used sensor data from 1,003,207 milkings from the University of Florida Dairy Unit collected from 2008 to 2014. Of these, 2,907 milkings were associated with recorded cases of clinical mastitis. Using the DLM/NBC method, we reached an area under the receiver operating characteristic curve of 0.89, with a specificity of 0.81 when the sensitivity was set at 0.80. Specifici-

ties with omissions of sensor data ranged from 0.58 to 0.81. These results are comparable to other studies, but differences in data quality, definitions of clinical mastitis, and time windows make comparisons across studies difficult. We found the DLM/NBC method to be a flexible method for combining multiple sensor and nonsensor data sources to predict clinical mastitis and accommodate missing observations. Further research is needed before practical implementation is possible. In particular, the performance of our method needs to be improved in the first 2 wk of lactation. The DLM method produces forecasts that are based on continuously estimated multivariate normal distributions, which makes forecasts and forecast errors easy to interpret, and new sensors can easily be added.

Key words: mastitis, Bayesian classifier, dynamic linear model

INTRODUCTION

Mastitis is associated with a wide range of characteristics that can be measured in milk. In a classic review, Kitchen (1981) described the effect of mastitis on the composition of milk and discussed potential diagnostics based upon these effects. In addition to SCC, electrical conductivity, milk constituents (especially lactose), and enzymes (such as *N*-acetyl- β -D-glucosaminidase and lactate dehydrogenase) have been identified to be affected by clinical mastitis.

Since the 1990s, work has been carried out on automated detection of mastitis using changes in one or more milk characteristics (e.g., Nielen et al., 1992). Automated mastitis detection systems started to be widely used on commercial dairy farms with the introduction of automatic milking systems approximately 20 yr ago. A mastitis detection system consists of at least 2 elements: the sensor (hardware) and the algorithms to translate sensor data into alerts (software). A decision support system and a decision-making system may also be part of a mastitis detection system (Rutten et

Received July 7, 2015.

Accepted May 1, 2016.

¹Corresponding author: daj@sund.ku.dk

al., 2013). The main sensor (hardware) used to detect mastitis measures electrical conductivity (e.g., Nielen et al., 1995b; Norberg et al., 2004; Caverio et al., 2006). Sensor systems based on other milk characteristics, such as milk color (Song et al., 2010), lactate dehydrogenase (Chagunda et al., 2006; Friggens et al., 2007), and SCC (Whyte et al., 2005; Mollenhorst et al., 2010), have also been proposed and are available on the market.

Most publications on automated mastitis detection systems are aimed at the algorithm for transforming sensor data into alerts. Significantly different data-modeling techniques have been proposed, including thresholds (e.g., Mollenhorst et al., 2010), moving averages (Maatje et al., 1992), neural networks (e.g., Nielen et al., 1995b; Caverio et al., 2008), fuzzy logic (e.g., de Mol and Woldt, 2001; Kamphuis et al., 2008), time series analysis (de Mol et al., 1999; Caverio et al., 2007), discriminant function analysis (Norberg et al., 2004; Kamphuis et al., 2010), and wavelet filtering (Mickleley et al., 2013). In most of these studies, electrical conductivity was combined with other measurements (mostly with milk yield) to improve the performance of the detection system.

So far, the performances of the published mastitis detection systems do not satisfy the high accuracy needed for practical clinical mastitis detection systems (Hogeveen et al., 2010). Combining data from more sources has been suggested as a possible method for improving the performance of mastitis detection systems. It remains unclear how to best combine data from different sensors and from other sources, including accounting for missing observations. Bayesian analysis has been used as an approach to prioritize sensor data-based alarms by including cow-specific information (Steenneveld et al., 2010).

Most mastitis detection systems compare observed sensor values to forecasted values and monitor forecast errors. Forecasts are typically based on moving averages (e.g., Maatje et al., 1992), but if the quality of the forecast is improved, then the performance of a mastitis detection system may be improved as well.

As a method for combining the many possible lines of sensor- and non-sensor-based data for a unified prediction of mastitis, we propose using a multivariate dynamic linear model (**DLM**) in combination with a naïve Bayesian classifier (**NBC**). The multivariate DLM provides the forecast values, whereas the NBC combines all available observations, including forecast errors, with a prior probability to achieve a single posterior probability of mastitis.

A property of the multivariate DLM, as described by West and Harrison (1997), is that it is adaptive, and thus the expected values are automatically adjusted to the longer term trend of the data. Another property

of the multivariate form of the DLM is that the codependencies between several variables of interest can be taken into account when one-step-ahead forecasts for these variables are calculated, which is attractive for the NBC.

Similar adaptive forecasting has been applied by Huybrechts et al. (2014), who used a synergistic control process to adjust lactation curves in an effort to use milk yield as a predictor of clinical mastitis (sensitivity: 0.63). Huybrechts et al. (2014) relied heavily on a specific mathematical model for long-term forecasting, whereas the adaptive and short-term nature of the forecasts produced by a DLM allows for a freer description of multiple (non)linear trends that may predict the short-term observations better. Furthermore, the DLM easily handles missing data because one-step-ahead forecasts are always produced given the available data.

Few applications of DLM for monitoring animal production systems exist. Univariate implementations of the DLM have been developed for applications including detection of estrus in sows (Ostersen et al., 2010) and describing the drinking behavior of young pigs (Madsen and Kristensen, 2005). To our knowledge, no previous descriptions of applications of a multivariate DLM exist for detecting diseases in production animals, such as mastitis in dairy cows.

An NBC classifies a new set of observations by estimating the probability that the observation belongs to each class (mastitis or healthy). The NBC is a relatively simple classification method, but it has been shown to be useful in a wide range of fields, such as prediction of bacterial thermophilicity (Jensen et al., 2012), diagnosis of classical swine fever (Geenen et al., 2011), and detection of clinical mastitis (Steenneveld et al., 2009). The NBC has advantages over comparable classification methods, such as artificial neural networks or logistic regression functions, because missing observations can be easily handled in an NBC by including only the observations that are available. Similarly, adding data from a new sensor is relatively trivial with the NBC, so long as likelihoods are available for the outputs of that sensor, associated with the outcome variable that needs to be classified. Such likelihoods may be estimated from scientific literature or practical knowledge of how mastitis influences milk characteristics and cow physiology, or they may be derived directly from observations made on site using the sensor. Lastly, the likelihoods make it easy to see the relative contributions of the various variables of interest, as opposed to the black-box nature of, for example, neural networks.

A combination of a DLM and NBC is therefore a potentially attractive practical method to detect clinical mastitis using data from multiple sources. This study had 2 objectives: (1) to describe and illustrate the

combination of a multivariate DLM and a NBC for detecting clinical mastitis, and (2) to measure the performance of the DLM/NBC method and estimate the relative importance of various combinations of sensor and nonsensor data on that performance.

MATERIALS AND METHODS

Data Sources

In this study, we refer to 2 types of data, namely, continuous data and categorical data. Continuous data are obtained with sensors and used by the DLM in their raw, numeric form. Categorical data are considered to fall within 2 or more separate categories, regardless of whether the information was collected using sensors.

All data were obtained from the University of Florida Dairy Unit (Hague, FL) between September 2008 and March 2014. The herd consisted of approximately 500 Holstein cows that were housed in freestall barns with sand bedding and fans and sprinklers for heat abatement. Cows were fed a TMR and milked twice per day with 12-h intervals in a double-12 milking parlor.

In and around the milking parlor, milk yield, milk conductivity, fat percentage, protein percentage, lactose percentage, blood percentage, SCC category, and BW were automatically recorded for each cow at each milking using sensors. Blood percentage is the volume of blood in 1 mL of milk expressed as a percentage. Milk yield and conductivity were measured with milk meters. Milk components and SCC categories were measured by real-time milk analyzers (AfiLab, Afimilk, Kibbutz Afikim, Israel). Body weights were measured by automated scales when cows exited the milking parlor. All sensors were obtained from Afimilk. The accuracies of the sensors for fat, protein, and lactose have been evaluated by Kaniyamattam and De Vries (2014). The 7 continuous variables from these sensor data were used in the DLM.

The 5 categorical variables were SCC category (0–200, 200–400, 400–800, > 800; \times 1,000 cells/mL), parity (first, later), previous mastitis treatment (yes, no, excluding treatment at the time of observation), season (warm: May to August; cold: September to April), and week in milk (**WIM**). These 5 categorical variables were not used in the DLM but were added in the NBC.

All records of clinical mastitis cases were obtained from the herd management information system. Clinical mastitis cases were determined in the milking parlor by trained farm staff following the farm's standard operating protocols (Donovan et al., 2011), which included forestripping and visual observation of milk. In addition, cows on alert lists for larger than expected milk

yield deviations were evaluated for clinical mastitis by farm staff.

If mastitis was diagnosed during the evening milking, the cow was immediately moved after her milking to the hospital herd (but not treated) and then evaluated and treated after her next milking in the morning. If mastitis was diagnosed during the morning milking, the cow was immediately moved to the hospital herd after her milking. The hospital herd was the last group milked in each milking shift. Treatment of mastitis occurred in the morning when the hospital herd was milked. Diagnosis and treatment were conducted by trained farm staff or veterinarians of the University of Florida. Confirmed clinical mastitis cases were treated according to the standard operating protocol and entered in the management information system.

All mastitis diagnoses were registered on a daily level, and the milking of first diagnosis was not available. For example, if mastitis was registered on a Friday, the actual mastitis diagnosis was made either during the Friday morning milking or the Thursday evening milking. Such daily mastitis recording is routine in the United States because the aim is to track withholding times for meat and milk after treatment.

Data Editing

All data editing, modeling, and calculations were done using the statistical language and environment R (The R Core Team, 2013). The sensor data were screened for outliers using simple summary statistics. We judged all observations from the 7 continuous sensor variables to be within acceptable ranges. Data collected for a given cow within at least 14 d of a new mastitis observation were not considered when calculating the updated probability of mastitis, and such data are not included in the following data descriptions.

All continuous sensor variables and SCC categories contained some missing values due to automatic data reading or entry problems in the parlor. The most extreme cases were BW and SCC category, with a total of 157,250 (15.67%) and 61,429 (6.12%) of these observations missing, respectively. The remaining sensor variables had between 15,595 (1.55%) and 31,071 (3.10%) missing observations. The missing data often occurred in the same milkings. The periods of missing data were between 1.32 and 1.84 observations long for the individual variables on average, but 9 out of the 2,051 lactations had periods of missing data for more than 100 consecutive observations for at least 1 sensor variable. The absolute highest number of consecutive missing observations was 257 observations ($n = 1$). None of the other categorical variables (i.e., season,

parity, previous mastitis treatments, and WIM) had missing observations.

New clinical mastitis cases were associated with both the morning and evening before milking because we could not distinguish if the case was first diagnosed in the evening or in the morning. Any case recorded within 14 d (28 milkings) of a new case was considered a flare up of the same case for that cow and was not counted separately.

The descriptive statistics of the edited data are shown in Table 1. Data from 1,003,207 milkings were available, including 2,907 milkings (morning and evening) from days on which clinical mastitis was recorded. Figure 1 shows the number and prevalence of mastitis observations in the first 43 WIM.

The available data were split evenly into a learning set (500,442 milkings, 1,455 of which were on days with mastitis diagnosis) for estimating parameters for the DLM and the likelihoods for the NBC, and a test set (502,765 milkings, 1,452 of which were on days with mastitis diagnosis) for validation of the method. This division was done by randomly assigning cows to the learning or test set and subsequently writing all observations from all lactations of the selected cows to their assigned set. This process ensured that the observations in the 2 sets were as independent of each other as possible.

Application of Multivariate DLM

Multivariate DLM with one-step Markov evolution (West and Harrison, 1997) was used to forecast the continuous sensor values for each individual cow at each milking. Because the mean level at any given DIM was clearly different between morning and evening milkings (Palmer et al., 1994) for several of the sensors, as indicated in Table 1, the morning and evening data were modeled separately. Thus, each DLM model would continuously forecast the 7 continuous sensor observations for the next morning or the next evening milking.

We expand on these methods because they are relatively unknown in the dairy science community. In general, a DLM consists of an observation equation and a system equation (Equations [1] and [2], respectively) as follows:

$$\mathbf{Y}_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(0, \mathbf{V}_t), \quad [1]$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(0, \mathbf{W}_t), \quad [2]$$

where \mathbf{Y}_t is the observation vector, \mathbf{F}_t' is the transposed design matrix, $\boldsymbol{\theta}_t$ is the unobservable parameter vector,

\mathbf{V}_t is the observational covariance matrix, \mathbf{G}_t is the system matrix, and \mathbf{W}_t is the systematic covariance matrix. Equation [1] describes how the values of an observation vector (\mathbf{Y}_t) depend on an unobservable parameter vector ($\boldsymbol{\theta}_t$) to time t . To describe both level and trend for each sensor, the parameter vector ($\boldsymbol{\theta}_t$) contains the underlying values for each of the continuous sensor variables, as well as the trend of the variable (i.e., the rates at which those same values change at time t). The initial mean levels for each variable were set as the mean value of the first observations of the respective variables across all lactations in the learning set. The initial trend was estimated as the average change in mean level from DIM = 1 to DIM = 2. These initial means and trends were estimate separately for morning and evening observations.

The system matrix (\mathbf{G}_t of Equation [2]) serves to update the expected values of the observable variables from time $t - 1$ to time t by adding the trend to the current level at that time. The transposed design matrix (\mathbf{F}_t') serves to extract the expected values of the observable variables from the parameter vector, thus yielding a vector that includes only these estimates.

In our case, the system matrix as well as the 2 variance matrices are constant, so that $\mathbf{G}_t = \mathbf{G}$, $\mathbf{V}_t = \mathbf{V}$, and $\mathbf{W}_t = \mathbf{W}$; \mathbf{F}_t varies over time because it depends on which of the 7 sensor variables have missing observations at a given time, as explained in the multivariate case in a following section.

Univariate Example

Assume that we wish to model the morning milk yield alone with a univariate DLM. We estimate that the initial morning milk yield on d 1 is 0.47 kg and that the initial trend in morning milk yield is +0.21 kg/d. Thus, we can describe the change in morning milk yield from DIM = 1 to DIM = 2 according to Equation [2] as follows:

$$\boldsymbol{\theta}_2 = \mathbf{G} \cdot \boldsymbol{\theta}_1 + \mathbf{w},$$

$$\boldsymbol{\theta}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.47 \\ 0.21 \end{bmatrix} + \mathbf{w},$$

$$\boldsymbol{\theta}_2 = \begin{bmatrix} 0.68 \\ 0.21 \end{bmatrix} + \mathbf{w}.$$

We can further describe the estimated milk yield observation at DIM = 2 according to Equation [1] as follows:

$$\mathbf{Y}_2 = \mathbf{F}_2' \cdot \boldsymbol{\theta}_2 + \mathbf{v},$$

Table 1. Descriptive statistics of the study data collected at the University of Florida Dairy Unit; data collected for a given cow within at least 28 milkings of a mastitis observation are not included

Categorical variable	N ¹	Mastitis cases	Continuous variables													
			Milk yield (kg)		Milk Conductivity (mS/cm)		Fat (%)		Protein (%)		Lactose (%)		Blood ² (%)		BW (kg)	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total	1,003,207	2,907	16.01	5.12	10.93	1.35	3.67	0.53	3.11	0.28	4.76	0.28	0.22	0.09	625.05	85.23
Time of day																
Morning	502,335	1,464	16.61	5.23	11.03	1.35	3.64	0.53	3.10	0.28	4.77	0.29	0.22	0.10	616.58	83.90
Evening	500,872	1,443	15.42	4.93	10.83	1.35	3.70	0.53	3.11	0.27	4.76	0.28	0.21	0.09	633.28	85.72
Current mastitis																
Yes	2,907	2,907	13.09	7.27	12.43	2.05	4.02	0.88	3.24	0.50	4.39	0.57	0.29	0.13	638.17	85.13
No	1,000,300	0	16.02	5.11	10.92	1.35	3.67	0.53	3.11	0.28	4.77	0.28	0.22	0.09	624.98	85.23
Season																
Cold	665,457	1,625	16.06	5.19	10.94	1.38	3.70	0.54	3.11	0.28	4.78	0.30	0.21	0.09	622.22	84.57
Warm	337,750	1,282	15.92	4.97	10.91	1.30	3.63	0.51	3.10	0.28	4.74	0.25	0.22	0.10	630.49	86.24
SCC category ³																
0-200	768,059	1,152	16.53	4.83	10.86	1.26	3.62	0.47	3.07	0.23	4.85	0.16	0.21	0.09	621.67	83.65
200-400	82,240	392	14.64	5.16	11.22	1.59	3.85	0.55	3.19	0.26	4.61	0.18	0.23	0.10	646.95	89.27
400-800	34,342	240	14.08	5.38	11.23	1.65	3.88	0.62	3.24	0.30	4.48	0.21	0.24	0.10	645.07	91.36
>800	57,137	563	13.48	5.66	11.24	1.65	3.88	0.73	3.32	0.40	4.21	0.31	0.27	0.11	638.85	91.58
Parity																
First	421,606	611	14.19	3.88	10.58	1.12	3.64	0.50	3.10	0.26	4.79	0.28	0.21	0.09	568.71	64.59
Later	580,396	2,314	17.32	5.50	11.18	1.45	3.70	0.55	3.11	0.29	4.75	0.29	0.22	0.10	664.89	74.96
Previous mastitis ⁴																
Yes	356,866	1,689	16.18	5.26	11.23	1.49	3.70	0.53	3.11	0.27	4.74	0.27	0.23	0.10	662.19	81.66
No	646,341	1,218	15.92	5.04	10.77	1.24	3.66	0.53	3.10	0.28	4.78	0.29	0.21	0.09	604.71	80.14

¹Number of milkings.

²Volume of blood in 1 mL of milk expressed as a percentage.

³In thousands of cells/mL.

⁴Binary variable, becomes “yes” after a cow has completed its first mastitis treatment.

$$\mathbf{Y}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.68 \\ 0.21 \end{bmatrix} + \mathbf{v},$$

$$\mathbf{Y}_2 = 0.68 + \mathbf{v}.$$

In other words, the prior estimates of the expected observations to a given time are calculated using the common rules for matrix multiplication.

The Multivariate Case

Constructing the multivariate DLM is accomplished by combining the univariate models needed to describe the individual variables, as shown in the example above, while taking the covariances between those variables into account. Here the transposed design matrix (\mathbf{F}'_t) and the system matrix (\mathbf{G}) have repeated structures corresponding to the number of variables being modeled, as illustrated by Equations [3] and [4], respectively.

$$\mathbf{F}'_t = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & & \vdots & \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad [3]$$

$$\mathbf{G}_t = \begin{bmatrix} 1 & 1 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \ddots & & \vdots & \\ 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & & 0 & 1 \end{bmatrix}. \quad [4]$$

When all 7 sensor variables are included, \mathbf{F}'_t has 7 rows and 14 columns, and \mathbf{G} has 14 rows and 14 columns. As seen in Equation [4], \mathbf{G} is a block-diagonal matrix in which each 2×2 block serves to update the expected value of a specific sensor variable by adding the appropriate trends to the current values, as in the univariate example described in the previous section. Also in accordance with the univariate example, the structure of \mathbf{F}'_t serves to separate those same observable values from the unobservable trends, yielding a vector of expected observations to each time t .

For \mathbf{F}'_t , only those rows corresponding to the variables that are actually observed at a given time are included at that time. This approach ensures that missing observations are ignored when the parameter vector is updated through Kalman filtering (West and Harrison, 1997). Kalman filtering is a method for reducing the noise in the observed data by considering the difference between the observed and forecasted values (i.e., the forecast error), while taking into consideration the variances associated with forecasting and observations.

The codependencies between the various observable variables (1 to n , with $n = 7$ in this study), as well as their unobservable trend values (d_1 to d_n) are accounted for by the observational covariance matrix (\mathbf{V}) and the system evolution covariance matrix (\mathbf{W}), as illustrated by Equations [5] and [6], respectively:

$$\mathbf{V} = \begin{bmatrix} V_{1,1} & \dots & V_{1,n} \\ \vdots & \ddots & \vdots \\ V_{n,1} & \dots & V_{n,n} \end{bmatrix}, \quad [5]$$

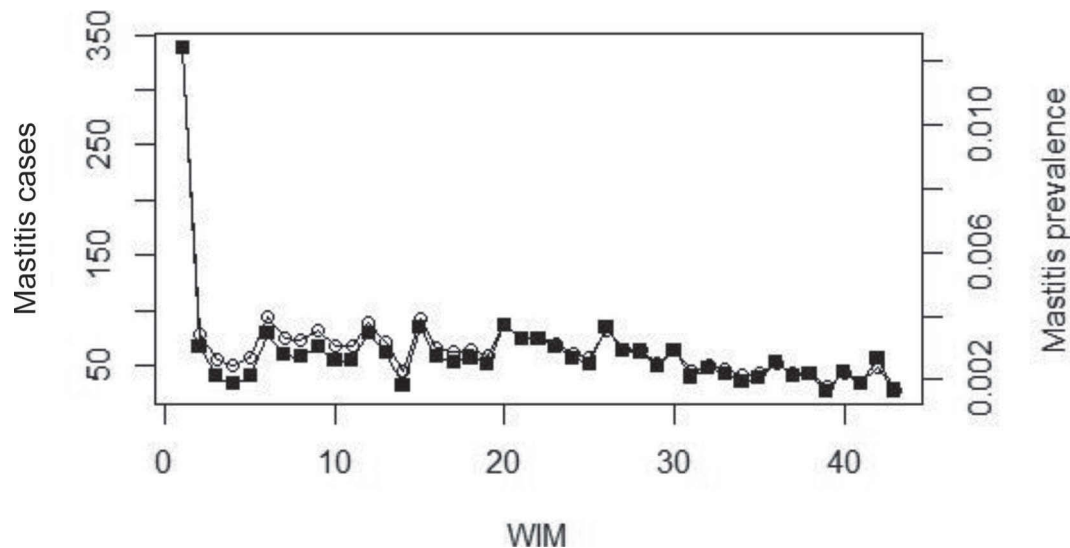


Figure 1. The number of mastitis cases (○; left vertical axis) and the prevalence of mastitis cases (■; right vertical axis) observed for the entire study population (i.e., all cows from the learning set and test set) by week in milk (WIM).

$$\mathbf{W} = \begin{bmatrix} W_{1,1} & W_{1,d1} & \cdots & W_{1,n} & W_{1,dn} \\ W_{d1,1} & W_{d1,d1} & \cdots & W_{d1,n} & W_{d1,dn} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ W_{n,1} & W_{n,d1} & \cdots & W_{n,n} & W_{n,dn} \\ W_{dn,1} & W_{dn,d1} & \cdots & W_{dn,n} & W_{dn,dn} \end{bmatrix}. \quad [6]$$

For example, if milk yield is considered variable number 1 and electrical conductivity is variable number 2, then $V_{1,1}$ is the observational variance of milk yield, $V_{2,2}$ is the observational variance of electrical conductivity, and $V_{1,2}$ is the observational covariance between milk yield and electrical conductivity. Similarly, $W_{1,1}$ is the systematic variance of the evolution of milk yield, $W_{1,d1}$ is the systematic covariance between milk yield evolution and the evolution of the trend (rate of change) of milk yield, and $W_{1,2}$ is the systematic covariance between the evolution of milk yield and electrical conductivity. Thus, the covariances found for the off-diagonal positions of the \mathbf{V} and \mathbf{W} matrices are what provides the extra information about variable interaction, compared with the information available in the univariate models.

We estimated the values in the \mathbf{V} and \mathbf{W} matrices using the expectation maximization (EM) algorithm (West and Harrison, 1997) applied to the learning set. The variances of the different variables were adjusted to similar scales by dividing all milk yield observations by 10 and all BW observations by 100 before modeling. To ensure that the DLM was optimized for modeling the healthy, nonmastitic state of the cows, only lactations with at least 43 full WIM, with at most 10% missing data overall and missing data periods of at most 2 consecutive observations, were used ($n = 41$ lactations) from the learning set. Convergence of the EM algorithm was determined by plotting variance values (i.e., the diagonal values of \mathbf{V} and \mathbf{W}) against the number of iterations and inspecting these plots visually. Convergence was reached when the EM algorithm was run for 300 iterations for the morning model and 230 iterations for the evening model.

To demonstrate how the multivariate DLM's ability to accurately forecast the continuous variables was affected by the presence of mastitis, the normalized forecast errors from the test set were selected for all morning milkings on days with and without mastitis, separately. For each continuous variable, the mean for nonmastitis forecast errors were calculated for each DIM. These healthy-associated mean errors were then plotted along with the individual forecast errors made for the same sensor variable on the days that the cow was positive for clinical mastitis.

Learning Likelihoods for Bayesian Classification

The forecast errors of a multivariate DLM are normally distributed, and any interdependencies between data from the 7 sensors are accounted for by the covariance matrices. The DLM was run from DIM = 1 to DIM = 301 for all lactations in the learning set. For each observation time (t) within a given lactation, a vector of forecast errors for all variables (\mathbf{e}_t) was generated. Simultaneously, a matrix describing the forecast covariances (\mathbf{Q}_t) was generated, as described by West and Harrison (1997). Using the standard deviations ($\sqrt{Q_t^i}$) derived from the forecast variances for each of the observed variables (i), the values of the forecast errors for the individual variables (e_t^i) were categorized in 1 of 4 observation categories (Obs_{*i*}): low ($e_t^i < 0 - \sqrt{Q_t^i}$), middle low ($0 > e_t^i > 0 - \sqrt{Q_t^i}$), middle high ($0 < e_t^i < 0 + \sqrt{Q_t^i}$), and high ($e_t^i > 0 + \sqrt{Q_t^i}$).

The likelihoods for each of the possible observation categories within each variable, given the possible 2 states [mastitis positive (clinical mastitis) or negative (healthy)], was calculated as

$$p(\text{Obs}_i|\text{Pos}) = N_{\text{Positive}}^{\text{Obs}_i} / N_{\text{Positive}}$$

and

$$p(\text{Obs}_i|\text{Neg}) = N_{\text{Negative}}^{\text{Obs}_i} / N_{\text{Negative}},$$

where $N_{\text{Positive}}^{\text{Obs}_i}$ is the number of occurrences of a particular observation category when the observed cow is known to be mastitis positive, and N_{Positive} is the total number of mastitis-positive observations. Similarly, $N_{\text{Negative}}^{\text{Obs}_i}$ is the number occurrences of the observation category when the observed cow is known to be mastitis negative, and N_{Negative} is the total number of mastitis-negative observations.

For each of the 5 categorical variables, the likelihood of observing each possible category given the 2 mastitis conditions (positive or negative) were calculated in the same way as for the categorized forecast errors from the 7 continuous sensor variables.

Application of Naïve Bayesian Classification

For each milking, the 5 categorical and 7 continuous variables were combined using an NBC, and a posterior probability of the cow being mastitis positive was calculated according to Bayes' formula in Equation [7]:

$$p(\text{Pos}|\text{Obs}_1, \dots, \text{Obs}_i) = \frac{\sum_{i=1}^{12} [p(\text{Obs}_i | \text{Pos})] \cdot p(\text{Pos})}{\sum_{i=1}^{12} [p(\text{Obs}_i | \text{Pos})] \cdot p(\text{Pos}) + \sum_{i=1}^{12} [p(\text{Obs}_i | \text{Neg})] \cdot p(\text{Neg})}, \quad [7]$$

where $p(\text{Obs}_i|\text{Pos})$ is the probability of the observation of the i th categorized variable given that the cow is mastitis positive, $p(\text{Obs}_i|\text{Neg})$ is the probability of the observation of the i th categorized variable given that the cow is mastitis negative, $p(\text{Pos})$ is the prior probability that the cow is mastitis positive, and $p(\text{Neg})$ is the prior probability that the cow is mastitis negative. In the learning set, we observed that $p(\text{Pos}) = 0.5\%$ and $p(\text{Neg}) = 99.5\%$.

If the observation of any variable was missing, then no likelihood related to that variable was included in the calculation of the posterior probability. The cow's milking was classified (predicted) as mastitis positive if the posterior probability was greater than a set threshold.

Performance Evaluation

All thresholds for positive classification between 0 and 1 with steps of 0.001 were evaluated. A posterior probability of mastitis above the set threshold was considered a mastitis alarm. For each threshold and milking, each mastitis alarm, or lack thereof, was compared with the diagnosis of mastitis provided by the farm staff.

In this study, the morning and evening data were modeled separately, and thus the performance of the DLM/NBC method was evaluated when applied separately to these 2 subsets. In this case, alarms were categorized as true positives (**TP**), false positives (**FP**), true negatives (**TN**), and false negatives (**FN**) on a single milking basis, depending on whether the farm staff had identified the cow as mastitis positive for that milking.

We also evaluated the performance achieved when combining the posterior probabilities produced by the 2 models before alarms were raised. Because all mastitis cases had either been observed in the morning on the day they were registered or in the previous evening, we allowed 3 ways in which a TP would be assigned to a single mastitis observation: (1) an alarm was raised based on both the morning probability and the previous evening probability, (2) an alarm was raised based only on the morning probability, and (3) an alarm was raised based only on the evening probability. Any of these 3 scenarios would count as 1 TP observation. Similarly, 1 FN observation was counted if no alarm was raised

based on either the morning probability or the evening probability for a given mastitis observation. The FP and TN were still assigned on a single milking basis. These definitions are illustrated in Figure 2.

This process resulted in lists of sensitivities (**Se**), calculated as $\text{Se} = \text{TP}/(\text{TP} + \text{FN})$, and specificities (**Sp**), calculated as $\text{Sp} = \text{TN}/(\text{TN} + \text{FP})$. The lists of Se and Sp, given the various thresholds, were used to plot receiver operating characteristics (**ROC**) curves (Zweig and Campbell, 1993), which were summarized by calculating the area under the curve (**AUC**) using the function auc from the library MESS in R.

In addition to the AUC of the ROC curve, the performances of the DLM/NBC method were evaluated based on the specificities achieved when sensitivity was held at 0.80, as recommended by Hogeveen et al. (2010), as well as the error rate of the predictions, calculated as $\text{Error rate} = (\text{FP} + \text{FN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN})$. The 95% CI for Se was calculated as $\text{Se}_{\text{CI}} = \text{Se} \pm \sqrt{(1 - \text{Se})/(\text{TP} + \text{FN})} \cdot 1.96$, and the 95% CI for Sp was calculated as $\text{Sp}_{\text{CI}} = \text{Sp} \pm \sqrt{(1 - \text{Sp})/(\text{TN} + \text{FP})} \cdot 1.96$.

Sensitivity Analysis

The sensors used in this study come as 3 separate sensor packages, each of which can be obtained independently of the others. Package 1 included milk meters that measured milk yield and electrical conductivity. Package 2 included the AfiLab that measured fat percentage, protein percentage, lactose percentage, blood percentage, and SCC category. Package 3 included the automatic scales that measured BW. We measured the performance of the DLM/NBC method achieved by each of the 7 possible combinations of the 3 sensor packages, as well as the effect of omitting the always-available nonsensor variables, which we refer to as package 0. These 8 combinations of information packages are summarized in Table 2. Statistically significant differences between the specificities of 2 information packages were identified by comparing their 95% CI.

RESULTS

DLM-Based Forecast Errors and Likelihoods

As is seen in Figure 3, milk conductivity, and to a lesser extent, fat percentage and blood percentage, showed tendencies toward positive forecast errors when mastitis occurred, and milk yield and lactose percentage showed tendencies toward being lower than forecasted by the DLM in mastitis-positive cows. From a purely visual inspection of these plots, the presence of mastitis did not notably affect the tendencies of protein

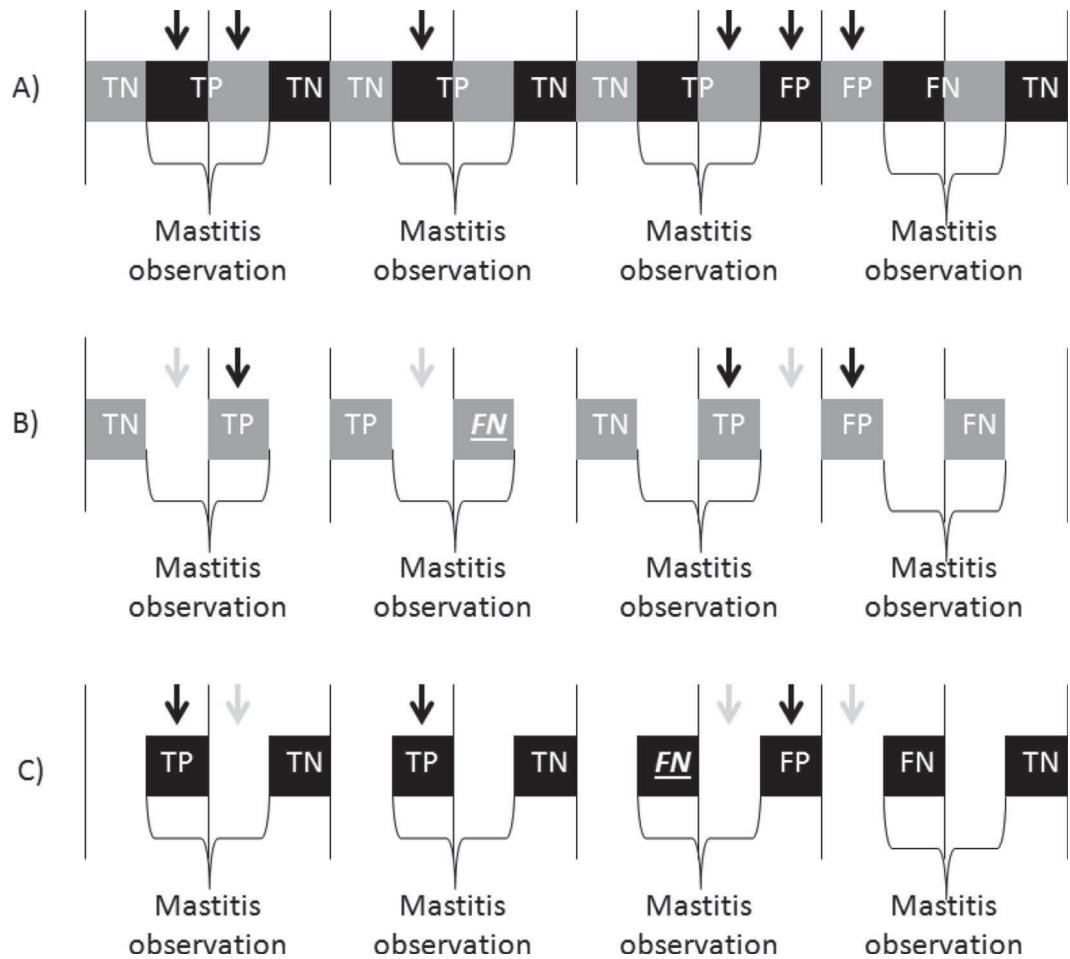


Figure 2. Illustration of the definitions of true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) alarms. Note that these examples serve only as illustrations and do not represent any actual observations. Gray blocks represent morning milkings, black blocks represent evening milkings, and thin vertical lines represent the separations between days. Mastitis observations are always associated with a morning milking and the preceding evening milking. Black arrows represent alarms raised by the considered models (morning, evening, or both); gray arrows represent alarms raised by the model, which are not considered (either morning or evening). (A) Alarms raised by the separate morning and evening models are both considered; (B) only alarms from the morning model are considered; (C) only alarms from the evening model are considered.

Table 2. The various combinations of the 4 information packages¹

Package combination	Included sensor packages	Included variables
1-2-3-0	Milk meter, AfiLab, automated scales, nonsensor information	Milk yield, conductivity, fat %, protein %, lactose %, blood %, ² SCC category, BW, parity, mastitis history, season, week in milk
1-2-3	Milk meter, AfiLab, automated scales	Milk yield, conductivity, fat %, protein %, lactose %, blood %, SCC category, BW
1-2-0	Milk meter, AfiLab, nonsensor information	Milk yield, conductivity, fat %, protein %, lactose %, blood %, SCC category, parity, mastitis history, season, week in milk
1-3-0	Milk meter, automated scales, nonsensor information	Milk yield, conductivity, BW, parity, mastitis history, season, week in milk
2-3-0	AfiLab, automated scales, nonsensor information	Fat %, protein %, lactose %, blood %, SCC category, BW, parity, mastitis history, season, week in milk
2-0	AfiLab, nonsensor information	Fat %, protein %, lactose %, blood %, SCC category, parity, mastitis history, season, week in milk
3-0	Automated scales, nonsensor information	BW, parity, mastitis history, season, week in milk

¹Package 1 = milk meter; package 2 = AfiLab (Afirmilk, Kibbutz Afikim, Israel); package 3 = automated scales; package 0 = nonsensor information.
²Volume of blood in 1 mL of milk expressed as a percentage.

percentage and BW toward being either above or below the forecasted value. Notice that even though the mastitis-associated forecast errors were not consistently found either above or below zero, they still provided useful information as long as they were more likely to be on one side than the other.

The likelihoods of the 5 categorical variables and the 7 categorized continuous variables are shown in Table 3. The likelihoods associated with both morning and evening milkings for the categorical variables were generally identical. For most continuous sensor data, the likelihoods differed considerably between morning and evening milkings.

Posterior Probabilities

Figure 4A shows the average posterior probabilities produced with the DLM/NBC method for lactations with no mastitis cases based on all available sensor and nonsensor data (Table 2, Package combination 1-2-3-0). Morning and evening milking probabilities are shown sequentially. In the beginning of the lactation, the DLM-informed probability of mastitis is fluctuating and generally high, before it settles to a more constant level of nearly 0. This pattern is typical because the DLM is adapting to the data of the individual cow during these first few days. Thus, an

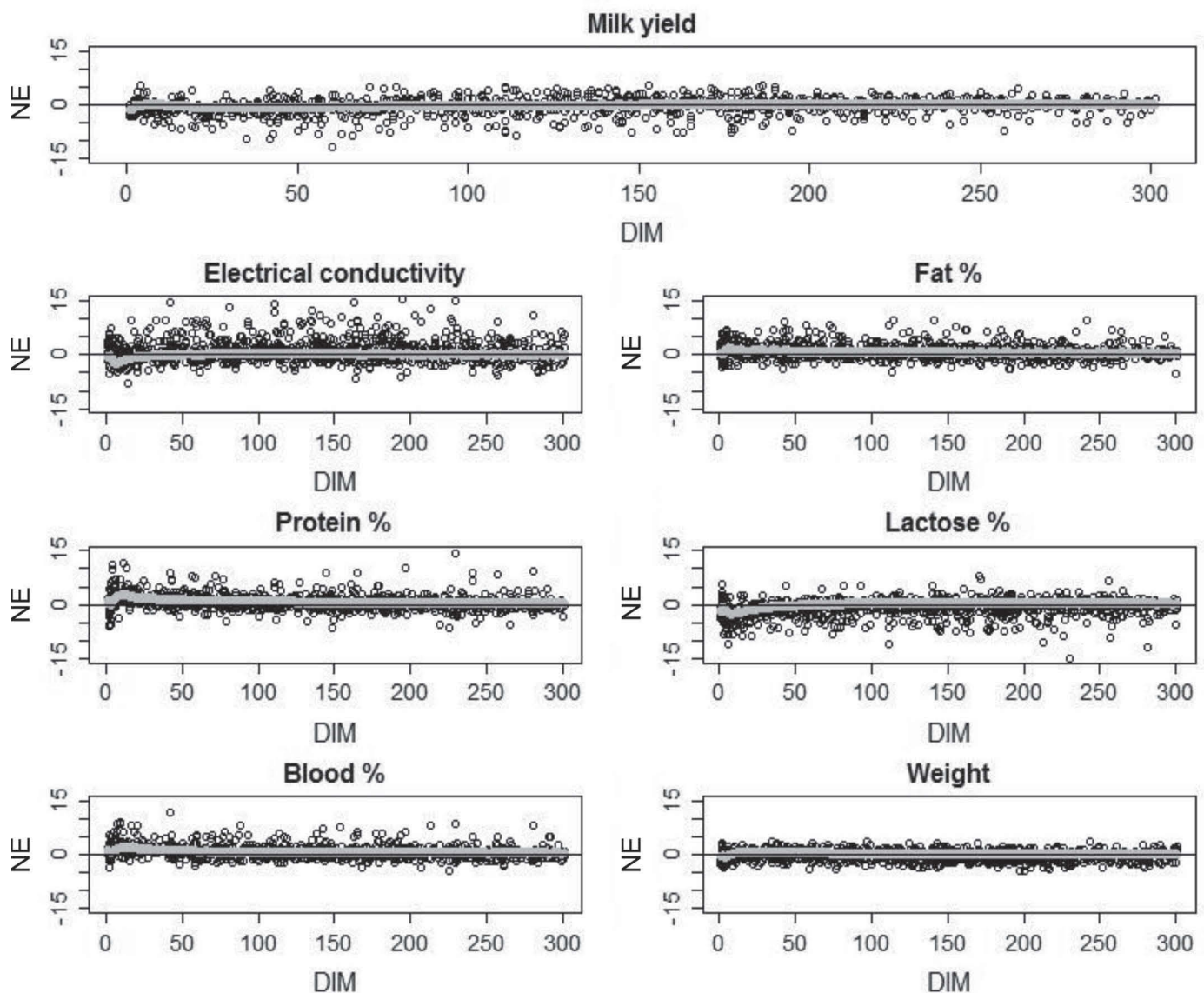


Figure 3. Normalized forecast errors given health state for each of the 7 continuous sensors. Points indicate errors when mastitis is observed; thick gray lines are the average errors when no mastitis is observed. The thin horizontal line marks the position of 0. Blood % = volume of blood in 1 mL of milk expressed as a percentage.

Table 3. Likelihood table for morning and evening milkings¹

Variable	Observation	Morning		Evening	
		$p(\text{Obs} \text{Pos})$	$p(\text{Obs} \text{Neg})$	$p(\text{Obs} \text{Pos})$	$p(\text{Obs} \text{Neg})$
SCC category ²	0–200	0.50	0.82	0.50	0.82
	200–400	0.15	0.09	0.15	0.09
	400–800	0.11	0.04	0.11	0.04
	>800	0.24	0.06	0.24	0.06
Previous mastitis	No	0.43	0.65	0.43	0.65
	Yes	0.57	0.35	0.57	0.35
Parity	Later	0.78	0.57	0.78	0.57
	First	0.22	0.43	0.22	0.43
Season	Cold	0.52	0.67	0.52	0.67
	Warm	0.48	0.33	0.48	0.33
WIM ³	1	0.12	0.03	0.12	0.03
	2	0.04	0.03	0.04	0.03
	3+	~0.02	~0.03	~0.02	~0.02
Milk yield	Low ⁴	0.31	0.09	0.31	0.21
	Middle low ⁵	0.37	0.59	0.32	0.43
	Middle high ⁶	0.23	0.30	0.21	0.24
	High ⁷	0.10	0.02	0.16	0.12
Conductivity	Low	0.24	0.20	0.27	0.21
	Middle low	0.20	0.37	0.19	0.36
	Middle high	0.17	0.32	0.17	0.29
	High	0.40	0.11	0.37	0.14
Fat %	Low	0.12	0.09	0.14	0.13
	Middle low	0.21	0.32	0.20	0.28
	Middle high	0.28	0.40	0.23	0.35
	High	0.39	0.19	0.43	0.25
Protein %	Low	0.18	0.13	0.13	0.10
	Middle low	0.25	0.34	0.21	0.23
	Middle high	0.28	0.37	0.29	0.37
	High	0.30	0.16	0.37	0.29
Lactose %	Low	0.38	0.16	0.69	0.59
	Middle low	0.29	0.36	0.17	0.27
	Middle high	0.22	0.38	0.06	0.07
	High	0.11	0.11	0.07	0.07
Blood % ⁸	Low	0.06	0.06	0.18	0.21
	Middle low	0.20	0.27	0.18	0.20
	Middle high	0.38	0.50	0.24	0.27
	High	0.36	0.18	0.4	0.32
BW	Low	0.47	0.31	0.29	0.15
	Middle low	0.26	0.31	0.22	0.22
	Middle high	0.18	0.25	0.24	0.28
	High	0.09	0.13	0.26	0.36

¹ $p(\text{Obs}|\text{Pos})$ is the likelihood of the observation given that the cow was diagnosed with mastitis; $p(\text{Obs}|\text{Neg})$ is the likelihood of the observation given that the cow was not diagnosed with mastitis.

²In thousands of cells/mL.

³Week in milk, from 1 to 43. After wk 3, the likelihood given mastitis condition is approximately constant.

⁴The observed value is >1 SD below the forecasted value.

⁵The observed value is <1 SD below the forecasted value.

⁶The observed value is <1 SD above the forecasted value.

⁷The observed value is >1 SD above the forecasted value.

⁸Volume of blood in 1 mL of milk expressed as a percentage.

elevated risk of FP exists in this stage of the lactation, which takes 15 d on average, as indicated by the vertical line.

Figure 4B shows a lactation in which a mastitis event is observed, marked by the dashed vertical line on DIM = 103. Notice the sharp spikes in the mastitis probability coinciding with this event, followed by at least 14 d during which the posterior probability is not

considered. Notice that spikes in the probability start to occur a few milkings before the event is observed. In this study, those premature spikes would be counted as FP, but they could conceivably be used for early warnings. Also, similarly to Figure 4A, fluctuations and higher-than-normal mastitis probabilities are seen at the beginning of the lactation, in spite of there being no observed mastitis event.

Predictive Performance

Table 4 shows the predictive performances of the 8 information package combinations described in Table 2.

Table 4 is sorted by the error rate, Sp, and AUC, which were achieved by combining the posterior probabilities produced by the morning and evening models before raising the alarms. In all cases, Se was held at 0.80 with

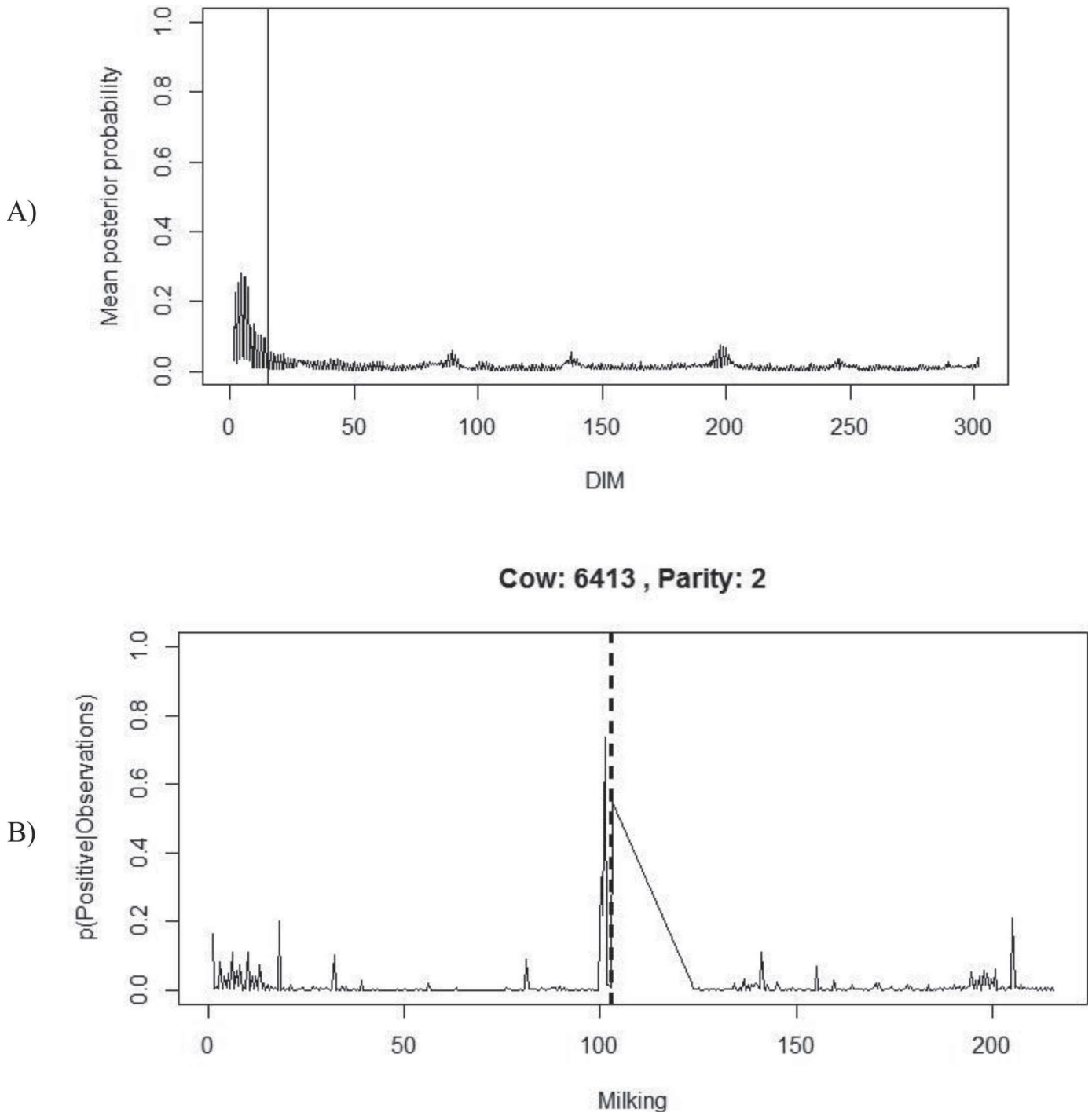


Figure 4. (A) The average posterior probability of mastitis given DIM for healthy lactations. The vertical line indicates DIM = 15; (B) posterior probabilities of mastitis for one cow during a single lactation. The thick, dashed vertical line at DIM = 103 indicates the observation of a mastitis event. Probabilities for morning and evening observations are plotted sequentially in both plots.

Table 4. The predictive performance of the 8 combinations of information packages as measured by the area under the receiver operating characteristic curve (AUC), specificity (Sp), and error rate (sensitivity, Se, was kept at 0.80)

Package combination ¹	Time of day	AUC	Threshold (Se = 0.8)	Sp	Sp 95% CI	Error rate
1-2-3-0	Combined	0.89	0.009	0.81	0.809–0.811	0.19
	Morning	0.85	0.007	0.75	0.748–0.752	0.25
	Evening	0.73	0.001	0.50	0.498–0.502	0.51
1-2-0	Combined	0.90	0.010	0.81	0.809–0.811	0.19
	Morning	0.84	0.008	0.75	0.748–0.752	0.26
	Evening	0.70	0.002	0.50	0.498–0.502	0.49
1-2-3	Combined	0.89	0.010	0.79	0.789–0.791	0.21
	Morning	0.85	0.006	0.68	0.678–0.682	0.33
	Evening	0.73	0.001	0.42	0.418–0.422	0.60
2-3-0	Combined	0.86	0.004	0.75	0.749–0.751	0.25
	Morning	0.82	0.004	0.64	0.638–0.642	0.35
	Evening	0.70	0.001	0.55	0.548–0.552	0.47
2-0	Combined	0.86	0.005	0.74	0.739–0.741	0.25
	Morning	0.81	0.004	0.63	0.628–0.632	0.37
	Evening	0.74	0.001	0.56	0.558–0.562	0.46
1-3-0	Combined	0.85	0.007	0.74	0.739–0.741	0.25
	Morning	0.79	0.004	0.65	0.648–0.652	0.34
	Evening	0.75	0.002	0.49	0.488–0.492	0.52
1-0	Combined	0.85	0.006	0.73	0.729–0.731	0.27
	Morning	0.78	0.004	0.63	0.628–0.632	0.38
	Evening	0.76	0.002	0.53	0.528–0.532	0.47
3-0	Combined	0.76	0.004	0.58	0.579–0.581	0.42
	Morning	0.73	0.004	0.53	0.528–0.532	0.46
	Evening	0.70	0.003	0.50	0.498–0.502	0.50

¹See Table 2 for package combinations.

the 95% CI being from 0.77 to 0.83 for the morning, the evening, and the combination-based alarms. The best performance of the DLM/NBC method using the full set of available sensor and nonsensor information (combination 1-2-3-0 in Table 2) resulted in an AUC of 0.89 with Sp of 0.81 and an error rate of 0.19.

Alarms that were raised based on the combination of the posterior probabilities from both morning and evening data consistently outperformed alarms that were raised based on the posterior probabilities from either model alone. This was true for all 3 performance measures. Similarly, the 3 performance measures consistently showed that alarms based on the morning model alone outperformed the alarms based on the evening model alone. These relationships are also evident from the ROC curves in Figure 5A.

From the ROC curves in Figure 5B, we see that the predictive performances achieved with the 8 information packages fall into 3 distinct groupings. The highest ROC curves are for the combinations 1-2-3-0, 1-2-0, and 1-2-3. In terms of all 3 performance measures, combinations 1-2-3-0 and 1-2-0 yielded exactly the same performances for the combination-based alarms, but the Sp achieved for combination 1-2-3 was significantly lower than with the other 2 package combinations at the 95% confidence level.

The second-highest group of ROC curves consists of the combinations 2-3-0, 1-3-0, 2-0, and 1-0. When Se was held at 0.80, the Sp and error rates achieved for combinations 3-1-0 and 2-0 were similar, although the AUC was slightly greater for the 3-1-0 combination (difference 0.01). Even though the AUC for the 1-0 combination was greater than that for the 2-3-0 combination, the 2-3-0 combination achieved preferable Sp and error rate values, when the Se was held at 0.80.

The third group consists solely of combination 3-0, showing the least favorable values with respect to all 3 performance measures.

DISCUSSION

Our main objective was to describe and demonstrate the combination of a multivariate DLM and an NBC as a novel but intuitive method for combining sensor and nonsensor data for detecting mastitis. We further assessed the performance that can be expected from this method when some sensor packages are not available.

DLM/NBC Methodology

The basic premise behind the NBC is that all included variables are mutually independent and are only

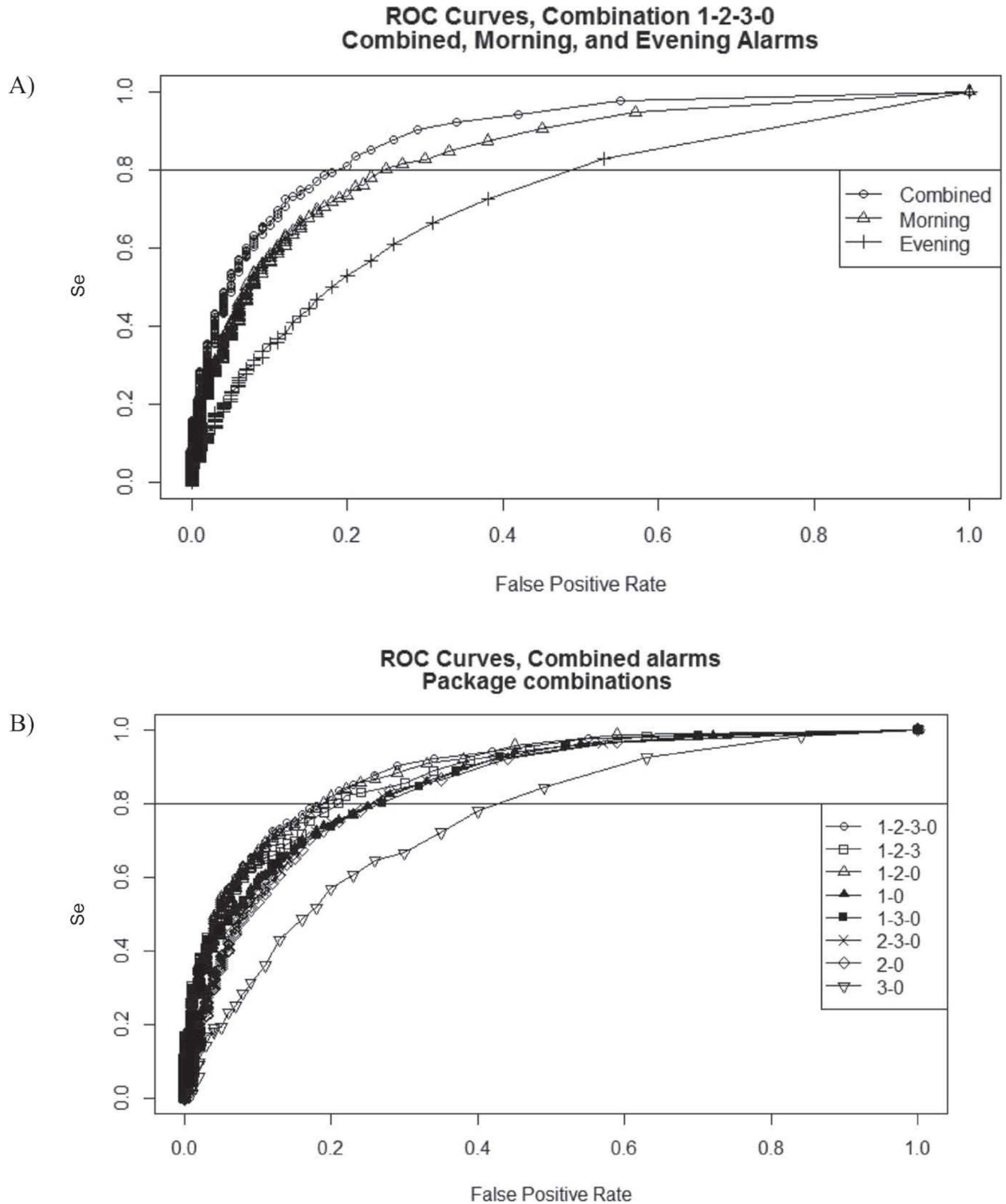


Figure 5. Receiver operating characteristic (ROC) curves for mastitis prediction: (A) ROC for combined alarms, morning alarms, and evening alarms, when all sensor and nonsensor information is included; (B) ROC for the combined alarms, given the 8 information package combinations included in this study (Table 2). Package 1 includes the milk meter, package 2 includes the AfiLab (AfiLab, Kibbutz Afikim, Israel), package 3 includes the automated scale, and package 0 includes nonsensor information.

affected by the outcome of interest. Even so, several of the sensor variables considered in this study, such as milk yield, the various milk components, and conductivity, are known to be highly correlated (Yoshida et al., 2005). The interdependencies between the 7 continuous variables are, however, accounted for by the act of co-modeling these variables in the multivariate DLM, taking into account the observational and systematic covariances between them, as described by the **V** and **W** matrices, respectively. It is worth noting that in situations in which the modeled variables are not subject to covariances, estimating and including such covariances does not make sense. In this study, we did not test whether the estimated co-variances were significant because that was considered to be beyond the scope of this study. No interdependencies between the 5 categorical variables are taken into account in our DLM/NBC method. For example, season as well as parity is known to affect milk composition (Yang et al., 2013). Nevertheless, the NBC method can yield excellent performances, even if the assumption of independence is violated (Pazzani, 1997). Furthermore, Steeneveld et al. (2010) attempted to improve their classification of TP mastitis cases by expanding their naïve Bayesian network to include dependencies between their included variables, but the resulting classification performance was not improved. We therefore consider the multivariate DLM/NBC method to be a reasonable compromise between accounting for codependencies between continuous variables while still allowing for easy incorporation of all available data, including the categorical nonsensor variables.

The higher risk of FP alarms during the first 15 DIM, as mentioned in relation to Figure 4, means that the performance during this period was worse than the overall performance (data not shown). Our algorithm cannot be very cow specific during this early period because not enough of the individual cow's data have been included yet. As seen on Figure 1, mastitis prevalence was at its highest during the first week of lactation, which makes accurate detection during this period particularly important. For this reason, more research needs to be done to improve the performance during this period before practical implementations are possible.

Table 3 shows that for some of the variables, the forecast errors related to the healthy cows did not follow an exact normal distribution with mean = 0 and 50% of the forecast errors on each side, as would generally be expected from a DLM. For milk yield, as an example, the forecast errors are generally biased toward the low and middle-low categories, but the opposite bias is seen for fat percentage in healthy cows. These deviations are generally caused by the simplistic way in which the lin-

ear trends in the DLM are estimated. A better but less simple way would be to estimate the trend based on some constant function; for example, the Wood curve (Wood, 1967), for the milk yield trend. In this case, the trend would always be defined by a function of the DIM multiplied by some factor learned by the DLM to adjust the curve to the level of the individual cow. Such functions should be implemented in future applications but were considered to be beyond the scope of this study. It is, however, important to recognize that mastitis detection is based on the differences in likelihoods given the 2 health states (mastitis or no mastitis). For this reason, it is less important that the forecast errors are not exactly normally distributed.

Performance of the DLM/NBC Method and Contribution from Sensor Packages

As seen in Table 4, the omission of the nonsensor information leads to a reduction in all 3 measures of predictive performance. This outcome is in agreement with the finding of Steeneveld et al. (2010) that including nonsensor information to distinguish between TP and FP mastitis alarms, raised by an automatic milking system, reduced the number of FP by 35%, thus increasing the specificity. Furthermore, although Steeneveld et al. (2010) included similar nonsensor data as were included in this study, they also included several other sources of nonsensor information that were not available for this study, and which they found to have considerably higher significance in distinguishing between TP and FP mastitis alarms. Thus, it stands to reason that even better performances could be achieved by including still more information.

Package combination 3-2-1-0 yielded the same performance measures as package combination 2-1-0, which suggests that if a farmer uses milk meters and the AfiLab, nothing more (in terms of mastitis detection) can be gained from investing in the automated scales. This finding does not mean that BW conveys no information, however. Its value is evident from the fact that package combination 2-3-0 performed better than 2-0 in terms of Sp and package combination 1-3-0 performed better than 1-0 in terms of Sp as well as error rate. As for package 1 compared with package 2, it is worth noting that the AUC is higher for package combination 1-0 compared with combination 2-0 (AUC = 0.860 vs. 0.848). One might thus be inclined to think that the milk meter provides better mastitis detection than the AfiLab, if one had to choose between having only one of these sensors. This conclusion is not the case, however, when Se is held at 0.80 because the corresponding Sp and error rate are both favorable for the AfiLab, with the difference in Sp being outside the 95% CI.

A clear cost–benefit analysis of investment in sensor packages was not feasible because real market prices were not available to us and sensors are usually also used for purposes other than mastitis detection. Routine maintenance time and costs were negligible for all sensors.

Table 4 shows that the evening data contain less useful information for mastitis detection than the morning data. This pattern is seen in, for example, the error rate for the evening model alarms generally being around 0.5 (0.46–0.60) when the sensitivity is held at 0.80. We speculate that this difference in information value is mainly because mastitis observations were only ever registered in the morning, even when they were made during the previous evening milking, thus making the mastitis labels more reliable for the morning part of the data. It is important to note that these error rates do not imply that the evening data contain no information. The value of these data is evident from the fact that the AUC values for the evening models are all well above 0.5 (0.70–0.76) as well as from the consistent increases in AUC when the morning and evening observations are combined, compared with only considering the posterior probabilities from the morning data. Table 4 also shows that the alarm thresholds needed to achieve Se of 0.8 with the evening models are generally very low (0.001–0.003). Consequently, to achieve Se of 0.8 with the evening models, one has to accept a high risk of FP alarms, which results in a high error rate. The thresholds for the morning as well as the combined posterior probabilities can be set higher while maintaining the same Se, which means fewer FP alarms and a lower error rate. Last, it should be noted that these seemingly very low threshold values are related to the prior probability of mastitis, which was estimated from the learning data to be 0.05. Had a flat prior been used, the thresholds would have been distributed around 0.50 in a similar fashion (data not shown). This possibility illustrates an important point about the NBC, namely that the classification of the observation (mastitis of no mastitis) should not be based on which class has the highest posterior probability, but rather on how the observations are able to change the prior probability in a positive or negative direction.

Results of the DLM/NBC Method Compared with the Literature

A direct comparison between our results and those reported in other studies does not fairly judge the performance of various methods, given that the different studies include data of different origins and unknowable quality. For example, case definitions of mastitis are not standardized and time windows of detection vary

(Hogeveen et al., 2010). We do, however, consider the following considerations to be worthwhile for comparison.

Some authors reported greater mastitis specificities than we found. For example, de Mol et al. (2001), Kamphuis et al. (2010), and Mollenhorst et al. (2010) presented mastitis detection with Sp of 0.979, 0.987, and 0.990, respectively. The associated Se were 0.67, 0.32, and 0.474, respectively, as compared with our set sensitivity of 0.80. Greater Sp results in lower Se. Figure 5 shows that higher Sp are obtainable at the cost of lower Se.

Other authors presented both high Se and Sp, but they typically used wider time windows in their performance evaluation, as opposed to the 2-milkings window used in our study when we combined the morning and evening alarms. For example, de Mol et al. (1997) used a time window of 17 d around the day of diagnosis (–10 d to +7 d, Se 0.90, Sp 0.982), and Cavero et al. (2006) used a 5-d window (–2 to +2, Se 0.929, Sp 0.939). Such long time windows may not be useful in practice. If an alarm is raised several days before clinical signs of mastitis are visible, the farmer will likely believe the alarm to be a FP. If this happens too frequently, the farmer will lose trust in the system, which would make it worthless (Hogeveen et al., 2010). Conversely, if an alarm is not raised until several days after clinical signs appear, treatment will be needlessly delayed. Thus, a very narrow time window is needed for practical mastitis detection applications.

We found only 3 studies with a combination of predictive performance and narrow time windows comparable to those presented in this paper. These studies presented mastitis detection based on artificial neural networks (Nielen et al., 1995a; Sun et al., 2010) and fuzzy logic (Kamphuis et al., 2008). These 3 studies were all conducted with relatively small study populations, which invariably reduces the reliability of any results. Specifically, Nielen et al. (1995a) included only 55 cows (31 with mastitis), and Kamphuis et al. (2008) included 18 mastitic cows. Given the reported number of cases and noncases in these studies, the 95% CI for the reported Se would include values as low as 0.71 and 0.59 for Nielen et al. (1995a) and Kamphuis et al. (2008), respectively. Sun et al. (2010) considered 194 cows, of which 43 (88 udder quarter milkings) were actually observed to be mastitic. However, Sun et al. (2010) inflated the number of infected quarter milkings to 895 by assuming that the quarters of the 43 cows were also mastitic at other times than indicated in the logbook, if the observed milk yield, conductivity, and SCC showed values above or below specific thresholds. They subsequently trained neural networks to detect mastitis based on observed values of milk yield and

conductivity, thus making the detection depend on the same variables that were used to define the majority of the events. Thus, the impressive performance ($Se = 0.87$, $Sp = 0.91$) shown by Sun et al. (2010) is likely due to incorporation bias.

A common approach in studies such as the preceding ones is to clear the data set of missing data before applying a detection method. This approach will invariably give an unfairly favorable impression of the performance that ultimately cannot be transferred to practice where missing data are unavoidable. In contrast, the DLM/NBC method is capable of handling missing observations, as was demonstrated with our realistic data sets in which all instances of missing data were preserved.

Perspectives

The performance of the DLM/NBC method might be improved if milk yield was corrected for the interval between milkings. Longer intervals are associated with greater milk yield (Palmer et al., 1994). The effects of other design choices, such as the selection of “healthy” lactations to estimate variance components, should also be further evaluated. The expectation maximization algorithm is time consuming, which can be a hindrance for the practical application of the DLM/NBC method. In this study, we chose to only include high-quality data when estimating the variance components, in part to reduce computation time. It would make sense to study the effect of the amount versus the quality of the data with respect to final model performance, which to our knowledge has not been done elsewhere. An alternative method could be to assume an unknown, non-constant system variance, which would be continuously estimated as described by West and Harrison (1997). This assumption of a nonconstant system variance might well be more biologically meaningful and should therefore be tested in future research. It is worth noting that even if the system variance was assumed to be nonconstant, the observational variance should still be considered constant because the sensors should not be expected to change through the course of the lactation or from one lactation to the next.

Given the performance demonstrated in this study, it is reasonable to investigate the method’s value for detecting other conditions in dairy cows or indeed health states in other production animals, such as pigs or poultry. Furthermore, one of the main advantages of using the NBC for information integration was the relative simplicity with which specific data could be ignored (e.g., in the case of missing data) or added to the probability calculation. A relevant follow-up study would be to investigate how much time is needed be-

fore information from a newly included sensor would add significantly to the performance of the DLM/NBC method, if the likelihoods need to be learned from observations in that herd. If the likelihoods of a condition associated with the values of the sensor can be directly applied between different herds, then a new sensor could potentially be useful immediately after its integration. However, if the likelihoods have to be estimated using on-site data, then the time before relevant information is added will depend heavily on the prevalence of the condition of interest.

CONCLUSIONS

A combination of a multivariate DLM to produce forecasts and an NBC using mastitis-dependent likelihoods of forecast errors can be meaningfully used to combine multiple types of data for detecting mastitis in dairy cows. An advantage of the proposed method is the ease with which missing observations can be handled and information from new sensors can be added, which are both necessary abilities in real-world farm situations. With this DLM/NBC method, we reached an AUC of 0.89, with a specificity of 0.81 when the sensitivity was held at 0.80 and when the alarms were raised based on the DLM forecast errors related to both morning and evening milkings. Although all sensor packages held some information relevant for mastitis detection, the automated scale was by far the least informative. The inclusion of nonsensor data significantly improved the performance. Further research is needed before this method is implemented in practice. This research should focus on better detection in the first weeks of lactation and the implementation of more biologically meaningful modeling of milk yield and milk composition.

ACKNOWLEDGMENTS

We thank the staff of the University of Florida Dairy Unit for their help with collecting the data for this study, in particular Eric Diepersloot and Grady Byers. Dan Jensen and Albert De Vries thank Henk Hogeveen for his hospitality during their stay at Wageningen University in November 2014. Dan Børge Jensen was financially supported by The Danish Council for Strategic Research (The PigIT project, grant number 11-116191). Albert De Vries received financial support from USDA-National Institute of Food and Agriculture Agriculture and Food Research Initiative grant award number 2013-68004-20424 entitled “Southeast Quality Milk Initiative: Controlling Mastitis and Improving Milk Quality.” We also thank veterinarian Line Sven-

neson for sharing her knowledge of the practical aspects of information availability on dairy farms.

REFERENCES

- Cavero, D., K. H. Tolle, C. Buxade, and J. Krieter. 2006. Mastitis detection in dairy cows by application of fuzzy logic. *Livest. Sci.* 105:207–213.
- Cavero, D., K. H. Tolle, C. Henze, C. Buxade, and J. Krieter. 2008. Mastitis detection in dairy cows by application of neural networks. *Livest. Sci.* 114:280–286.
- Cavero, D., K. H. Tolle, G. Rave, C. Buxade, and J. Krieter. 2007. Analysing serial data for mastitis detection by means of local regression. *Livest. Sci.* 110:101–110.
- Chagunda, M. G. G., N. C. Friggens, M. D. Rasmussen, and T. Larsen. 2006. A model for detection of individual cow mastitis based on an indicator measured in milk. *J. Dairy Sci.* 89:2980–2998.
- de Mol, R. M., A. Keen, G. H. Kroeze, and J. Achten. 1999. Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis combined with a Kalman filter. *Comput. Electron. Agric.* 22:171–185.
- de Mol, R. M., G. H. Kroeze, J. M. F. H. Achten, K. Maatje, and W. Rossing. 1997. Results of a multivariate approach to automated oestrus and mastitis detection. *Livest. Prod. Sci.* 48:219–227.
- de Mol, R. M., W. Ouweltjes, G. H. Kroeze, and M. M. W. B. Hendriks. 2001. Detection of estrus and mastitis: Field performance of a model. *Appl. Eng. Agric.* 17:399–407.
- de Mol, R. M., and W. E. Woldt. 2001. Application of fuzzy logic in automated cow status monitoring. *J. Dairy Sci.* 84:400–410.
- Donovan, A., C. Risco, F. Manusell, J. Lemmerman, and E. Diepersloot. 2011. IFAS Dairy Unit Standard Operating Procedures, Univ. Florida, Gainesville. Accessed May 31, 2016. <http://animal.ifas.ufl.edu/facilities/du/FloridaDairyUnitSOP11-07-01.pdf>.
- Friggens, N. C., M. G. G. Chagunda, M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen. 2007. Estimating degree of mastitis from time-series measurements in milk: A test of a model based on lactate dehydrogenase measurements. *J. Dairy Sci.* 90:5415–5427.
- Geenen, P. L., L. C. van der Gaag, W. L. A. Loeffen, and A. R. W. Elbers. 2011. Constructing naïve Bayesian classifiers for veterinary medicine: A case study in the clinical diagnosis of classical swine fever. *Res. Vet. Sci.* 91:64–70.
- Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010. Sensors and clinical mastitis—The quest for the perfect alert. *Sensors (Basel)* 10:7991–8009.
- Huybrechts, T., K. Mertens, J. De Baerdemaeker, B. De Ketelaere, and W. Saeys. 2014. Early warnings from automatic milk yield monitoring with online synergistic control. *J. Dairy Sci.* 97:3371–3381.
- Jensen, D. B., T. C. Vesth, P. F. Hallin, A. G. Pedersen, and D. W. Ussery. 2012. Bayesian prediction of bacterial growth temperature range based on genome sequences. *BMC Genomics* 13(Suppl 7):S3.
- Kamphuis, C., H. Mollenhorst, A. Feelders, D. Pietersma, and H. Hogeveen. 2010. Decision-tree induction to detect clinical mastitis with automatic milking. *Comput. Electron. Agric.* 70:60–68.
- Kamphuis, C., R. Sherlock, J. Jago, G. Mein, and H. Hogeveen. 2008. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. *J. Dairy Sci.* 91:4560–4570.
- Kaniyamattam, K., and A. De Vries. 2014. Agreement between milk fat, protein, and lactose observations collected from the Dairy Herd Improvement Association (DHIA) and a real-time milk analyzer. *J. Dairy Sci.* 97:2896–2908.
- Kitchen, B. J. 1981. Review of the progress of dairy science: Bovine mastitis: Milk compositional changes and related diagnostic tests. *J. Dairy Res.* 48:167–188.
- Maatje, K., P. J. M. Huijsmans, W. Rossing, and P. H. Hogewerf. 1992. The efficacy of in-line measurement of quarter milk electrical conductivity, milk yield and milk temperature for the detection of clinical and subclinical mastitis. *Livest. Prod. Sci.* 30:239–249.
- Madsen, T. N., and A. R. Kristensen. 2005. A model for monitoring the condition of young pigs by their drinking behaviour. *Comput. Electron. Agric.* 48:138–154.
- Miekley, B., E. Stamer, I. Traulsen, and J. Krieter. 2013. Implementation of multivariate cumulative sum control charts in mastitis and lameness monitoring. *J. Dairy Sci.* 96:5723–5733.
- Mollenhorst, H., P. P. J. van der Tol, and H. Hogeveen. 2010. Somatic cell count assessment at the quarter or cow milking level. *J. Dairy Sci.* 93:3358–3364.
- Nielen, M., H. Deluyker, Y. H. Schukken, and A. Brand. 1992. Electrical conductivity of milk: Measurement, modifiers and meta analysis of mastitis detection performance. *J. Dairy Sci.* 75:606–614.
- Nielen, M., Y. H. Schukken, A. Brand, S. Haring, and R. T. Ferwerdavan Zonneveld. 1995a. Comparison of analysis techniques for on-line detection of clinical mastitis. *J. Dairy Sci.* 78:1050–1061.
- Nielen, M., M. H. Spigt, Y. H. Schukken, H. A. Deluyker, K. Maatje, and A. Brand. 1995b. Application of a neural network to analyze online milking parlor data for the detection of clinical mastitis in dairy cows. *Prev. Vet. Med.* 22:15–28.
- Norberg, E., H. Hogeveen, I. R. Korsgaard, N. C. Friggens, K. H. M. N. Sloth, and P. Lovendahl. 2004. Electrical conductivity of milk: Ability to predict mastitis status. *J. Dairy Sci.* 87:1099–1107.
- Ostersen, T., C. Cornou, and A. R. Kristensen. 2010. Detecting oestrus by monitoring sows' visits to a boar. *Comput. Electron. Agric.* 74:51–58.
- Palmer, R. W., E. L. Jensen, and A. R. Hardie. 1994. Removal of within-cow differences between morning and evening milk yields. *J. Dairy Sci.* 77:2663–2670.
- Pazzani, P. D. M. 1997. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Mach. Learn.* 29:103–130.
- Rutten, D. J., G. J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: Sensors to support health management on dairy farms. *J. Dairy Sci.* 96:1928–1952.
- Song, X., S. Zhuang, and P. P. J. van der Tol. 2010. New model to detect clinical mastitis in Astronaut A3 next milking robot. Page 2689–2693 in *Mastitis Research into Practice*. J. E. Hillerton, ed. VetLearn, Wellington, New Zealand.
- Steeneveld, W., L. C. van der Gaag, H. W. Barkema, and H. Hogeveen. 2009. Providing probability distributions for the causal pathogen of clinical mastitis using naïve Bayesian networks. *J. Dairy Sci.* 92:2598–2609.
- Steeneveld, W., L. C. van der Gaag, W. Ouweltjes, H. Mollenhorst, and H. Hogeveen. 2010. Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems. *J. Dairy Sci.* 93:2559–2568.
- Sun, Z., S. Samarasinghe, and J. Jago. 2010. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. *J. Dairy Res.* 77:168–175.
- The R Core Team. 2013. R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- West, M., and J. Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. 2nd ed. Springer, New York, NY.
- Whyte, D., M. Walmsley, A. Liew, R. Claycomb, and G. Mein. 2005. Chemical and rheological aspects of gel formation in the California Mastitis Test. *J. Dairy Res.* 72:115–121.
- Wood, P. D. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164–165.
- Yang, L., Q. Yang, M. Yi, Z. H. Pang, and B. H. Xiong. 2013. Effects of seasonal change and parity on raw milk composition and related indices in Chinese Holstein cows in northern China. *J. Dairy Sci.* 96:6863–6869.
- Yoshida, T., N. Lopez-Villalobos, and C. W. Holmes. 2005. Relationships between milk yield, milk composition and electrical conductivity in dairy cattle. *Proc. N.Z. Soc. Anim. Prod.* 65:143–147.
- Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39:561–577.